DOCUMENT RESUME

ED 159 175 95 TM 007 317

AUTHOR Madey, Loren L.
TITLE The Evaluator's Perspective: Evaluating the State Capacity Building Program.
INSTITUTION National Testing Service, Inc., Durham, N.C.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Feb 78
CONTRACT 400-76-0166
NOTE 49p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Cost Effectiveness; Data Analysis; Data Collection; *Evaluation Methods; Federal Programs; Grants; *Information Dissemination; Information Networks; *Methods Research; Program Evaluation; *Research Design; Research Methodology; State Departments of Education; *State Programs
IDENTIFIERS *Qualitative Research; *Quantitative Research

ABSTRACT

A historical antagonism between the advocates of quantitative evaluation methods and the proponents of qualitative evaluation methods has stymied the recognition of the value to be gained by utilizing both methodologies in the same study. The integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. An ongoing study, the evaluation of the State Capacity Building Grants Program in educational dissemination, is used to illustrate the benefits obtained by integrating methods during the design, data collection, and analysis phases. The examples underscore that timing considerations are of major importance to the management of an evaluation which seeks to benefit from both methodologies. Three time sequences are possible: qualitative methods can precede, follow, or proceed concurrently with quantitative methods. The costs involved in such a methodological marriage are also explored. For longitudinal evaluations, the benefits seem to outweigh the costs. The integration of methods offers enormous potential for improving evaluation strategies. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Evaluator's Perspective:
Evaluating the State Capacity Building Program

ED159175

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Doren L. Madey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

Doren L. Madey

NTS Research Corp.
Durham, North Carolina

TM007 317

February 1978

We must use all available weapons of attack, face our problems realistically and not retreat to the land of fashionable sterility, learn to sweat over our data with an admixture of judgment and intuitive rumination, and accept the usefulness of particular data even when the level of analysis available for them is markedly below that available for other data in the empirical area.

Binder, Statistical Theory

The work upon which this paper is based was performed pursuant to Contract 400-76-0166 with the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. Contractors undertaking such projects with government sponsorship are encouraged to express their professional judgment in the conduct of the project. Points of view or opinions do not, therefore, necessarily represent official positions or policies of NIE, and no official endorsement by NIE should be inferred.

Printed in U.S.A.

TABLE OF CONTENTS

	Page
Figures and Tables	ii
Abstract	1
I. Introduction	1
II. The State Capacity Building Program and Its Evaluation	4
III. Benefits Accrued in Integrating Methods	14
A. Contributions of Qualitative Methods to Quantitative Methods	14
B. Contributions of Quantitative Methods to Qualitative Methods	22
IV. Costs Involved in Integrating Methods	28
V. Applicability to Other Evaluations	31
Footnotes	33
References	35
Appendix A. Evaluation Questions for the NTS Study	39

LIST OF FIGURES

Figure 1. States Receiving Grants through NIE's State Capacity Building Program (SCBP) 7

Figure 2. Twelve Activities Contributing to Final Design Specifications 10

LIST OF TABLES

Table 1. Respondent Universe and Estimated Sample Size for the First Year of the Evaluation 11

Table 2. Client Samples for States in the SCBP 12

THE EVALUATOR'S PERSPECTIVE:

EVALUATING THE STATE CAPACITY BUILDING PROGRAM

Doren L. Madey
NTS Research Corporation

A historical antagonism between the advocates of quantitative evaluation methods and the proponents of qualitative evaluation methods has stymied the recognition of the value to be gained by utilizing both methodologies in the same study. The integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. An ongoing study, the Evaluation of the State Capacity Building Grants Program in Dissemination, is used to illustrate the benefits obtained by integrating methods during the design, data collection, and analysis phases. The examples underscore that timing considerations are of major importance to the management of an evaluation which seeks to benefit from both methodologies. The costs involved in such a methodological marriage are also explored. For longitudinal program evaluations, the benefits seem to outweigh the costs. Integrating methods offers enormous potential for improving evaluation strategies.

I. INTRODUCTION

Prior to this decade, tightly controlled experimental designs dominated the evaluation field. Classical Fisherian experiments,[1] preferably using factorial designs, were extolled as the most powerful and acceptable model for evaluation (Fisher, 1935 et. seq.; Boring, 1954). Where the stringent requirements of the tightly controlled experimental design could not be met, quasi-experimental designs were considered viable alternatives (Campbell and Stanley, 1966). Program evaluators recognized that to follow the experimental model faithfully would be too difficult to design and to implement (Evans, 1969). But qualitative designs were shunned as too imprecise to produce either information of value to decisionmakers or information to establish program accountability (Rossi, 1972):

> There is almost universal agreement among evaluation researchers that the randomized controlled experiment is the ideal model for evaluating the effectiveness of public policy. If there is a Bible for evaluation, the Scriptures have been written by Campbell and Stanley....[2]

Following release of preliminary results of a Westinghouse Learning Corporation-Ohio University evaluation, which indicated that "the long term effect of Head Start appears to be extremely weak,"[3] a major controversy raged on how best to evaluate federal programs. The Head Start Evaluation had employed a quasi-experimental (ex post facto) design, which, until the results were released, had been heralded as yielding the best evaluation information one could hope to obtain (Rossi, 1972). With repetitive announcements of similar "weak" results by other evaluations, decisionmakers began to realize that knowing whether a program worked or did not work would not be enough. For program improvement purposes, decisionmakers would also need to know why a program did or did not work (Mason, 1973).

Dissatisfaction with the vogue methodologies resulted in a search for viable alternatives. Studies employing qualitative methodologies emerged as a solution (Sieber, Louis, and Metzger, 1972; Murphy, 1974; Partlett and Hamilton, 1975; Stake, 1978; Wolf and Tymetz, n.d.):

> Some people assume that natural inquiry is unstructured, undisciplined, and unsystematic; let it suffice to say that such thoughts are unfair, inaccurate, and simplistic. Natural inquiry is no less rigorous than traditional experimental research, it is simply different. The paradigm of natural inquiry is comprehensive in scope, demanding in design, and requires a set of honorable skills that even some rigid experimentalists lack (although they might be embarrassed to admit it).[4]

In this decade, qualitative evaluation designs for federal programs are beginning to achieve the respect once reserved only for classical experimental designs, but the controversy on how best to evaluate federal programs rages on.

Two warring camps are firmly entrenched in a methodological battle.[5] One camp extols the virtues of hard, generalizable data (Campbell and Stanley, 1966) and the other superiority of deep, rich observational data (Stake, 1978). In principle, evaluators may agree on the appropriateness of combining methods, but in practice, evaluators seldom integrate methods. Instead they tend to employ the methods of one of the two camps. Reasons for concentrating on one methodology include limited time, money, and skills. The antagonism between the two camps has stymied the actual utilization of

both quantitative and qualitative methods in one study. Actual examples of "integrated evaluations" are scarce.

The polarization has also hidden the fact that a continuum of methods actually exists.[6] Another continuum, the color spectrum, provides an analogy. Just as adding blue to the red end of the spectrum deepens the hue, so using a mixture of evaluation methods enriches the information yield. In this paper, the focus is not on describing the gradations of the continuum. Rather, this paper examines the ends of the continuum to show how integrating methods can enrich and improve evaluation design, data collection, and analysis.

Just as a standard definition of evaluation does not exist (cf. Stenner, 1974; Ross and Cronbach, 1976), the terms "quantitative" and "qualitative" have not been uniformly demarcated. For the purposes of this paper, evaluation is defined as a "process of clarifying a set of information needs, and collecting, analyzing, and reporting information to alleviate needs." (Stenner, 1974, p. 4). Qualitative methods are defined as relatively open-ended methods for collecting data. They include "the more episodic, subjective procedures, common to the case study" (Stake, 1978, p. 6) and other "soft" methods of "naturalistic inquiry" (Guba, 1978, in press). Qualitative methods range from observation and other unobtrusive forms of interaction in natural settings (Webb et al., 1966) to open-ended probes, ethnographic techniques, case studies, and structured interviews. Quantitative methods are defined as relatively closed-ended methods for collecting data. Included are survey instruments, closed-ended questionnaires, indices and scales, achievement and aptitude tests, and randomized, controlled experiments (Stanley and Campbell, 1966; Cooley and Lohnes, 1976).

One basic distinction between qualitative and quantitative methods is the stage of evolution of the constructs. Often in qualitative studies, constructs are less evolved than in quantitative studies. Constructs in a qualitative study are identified and developed using "soft" methods of data collection. In a quantitative study, the constructs generally already exist in a relatively refined state. Well-defined constructs can be measured using quantitative data collection methods. For example, in an evaluation of a reading program the outcome constructs

to be measured already exist (e.g., reading achievement, vocabulary, comprehension) and quantitative methods are quite useful in such instances. But to understand the process of reading (e.g., how a child learns vocabulary, how a child increases his/her reading comprehension), more qualitative methods are needed as these process constructs are less refined.

The major thesis of this paper is that the integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. Integrating methods is a viable mechanism for designing and implementing a "collaborative evaluation." An ongoing study, the Evaluation of the State Capacity Building Grants Program in Dissemination, is used to illustrate the benefits obtained by integrating methods. The paper deals first with the contributions of qualitative methods to quantitative work using examples from the Evaluation of the State Capacity Building Grants Program. Then, the sequence is reversed and the contributions of quantitative to qualitative methods are examined. Once again, illustrations from the ongoing evaluation are used to highlight the synergistic effects. The examples underscore the view that timing considerations are of major importance to the management of an evaluation which seeks to benefit from both methodologies. A separate section explores the costs involved in such a methodological marriage. The final section explores the applicability of integrated methodologies to other evaluations.

II. THE STATE CAPACITY BUILDING PROGRAM AND ITS EVALUATION

Prior to discussing the benefits and costs associated with an evaluation which attempts to integrate methodologies, it is important to describe the specific program and evaluation which will be used to illustrate the integration of methods. This section presents a brief overview of the State Capacity Building Program and then describes the evaluation--its purposes, activities, respondent groups, and instrumentation.

The Program

As part of its mission to foster improvement of education throughout the country, the National Institute of Education (NIE) is responsible for building "an effective educational research and development system." The State Dissemination Grants Program is a major initiative of NIE to achieve its mission. The Program is also part of a configuration of federally supported efforts to increase the quantity and quality of dissemination of educational knowledge, programs, and practices. The State Dissemination Grants Program is designed to improve educational practice by encouraging State Education Agencies (SEAs) to develop their capacity for gaining access to information resources and to link such information to the needs of educators. The Program is administered by the Information and Communications Branch of the Dissemination and Resources Group in NIE.

Through the State Dissemination Grants Program, two types of awards are made to SEAs:

- Capacity Building Grants support SEA efforts to build a comprehensive state dissemination capacity. These awards, averaging about $100,000 each, are of one-year duration and potentially renewable for three-to-five years.

- Special Purpose Grants, averaging $25,000 each, support relatively low-cost, short-term SEA improvement efforts related to building a comprehensive state dissemination capacity. These grants are used to support such SEA activities as initial planning, training of personnel, or the development of specific dissemination resources.

Through the State Dissemination Grants Program, in 1975, ten states were awarded Capacity Building Grants and four states received Special Purpose Grants. In 1976, fourteen Capacity Building Grants and five Special Purpose Grants were awarded. In Fiscal Year 1978, six additional Capacity Building Grants and at least four more Special Purpose Grants will be awarded. Pending availability of funds, NIE anticipates supporting successive groups of states until all those qualified are included.

The State Capacity Building Program is comprised of those states which have received Capacity Building Grants. Figure 1 indicates those SEAs which currently have NIE grants to build dissemination capacity.

The State Capacity Building Program (SCBP) has been defined by broad NIE goals. The Grants Announcements state that establishing a comprehensive dissemination capacity requires three components: a comprehensive set of resources, a means of linking client groups to the resource base, and the leadership and management arrangements which facilitate the provision of services on any problem to all clients. The Grants Announcements also state that the Program is designed to be user-driven, that is, to respond to any information request from any client. SEAs have been allowed without specific prescription to develop their own objectives from these broad goals; thus the program has been called "non-prescriptive." Timelines are mutually agreed upon by NIE and the states as part of the collaborative effort.

The Evaluation

Under the sponsorship of Dissemination and Resources Group (DRG), NTS is conducting a two-phased study of the State Capacity Building Grants Program. NTS is currently completing an extended seventeen-month design phase; the full-scale evaluation is scheduled to begin in April 1978 and will extend until six months after the last group of states completes its funding period. This means that the NTS Evaluation is a longitudinal evaluation, which will continue, pending availability of funds, through FY 1982 even if no additional states are funded.

The NTS Study is designed to evaluate the process of developing a comprehensive dissemination capacity in SEAs and to assess the Program's impact on the projects, the SEAs, and the states. The Special Purpose Grants are not included in the NTS Evaluation since they are only of one year's duration and are not intended to be comprehensive in scope.

In the NTS Study, as in the Grants Announcements which describe the Program, dissemination is defined as "a two-way process for communicating knowledge relevant to educational needs and problems so that educational

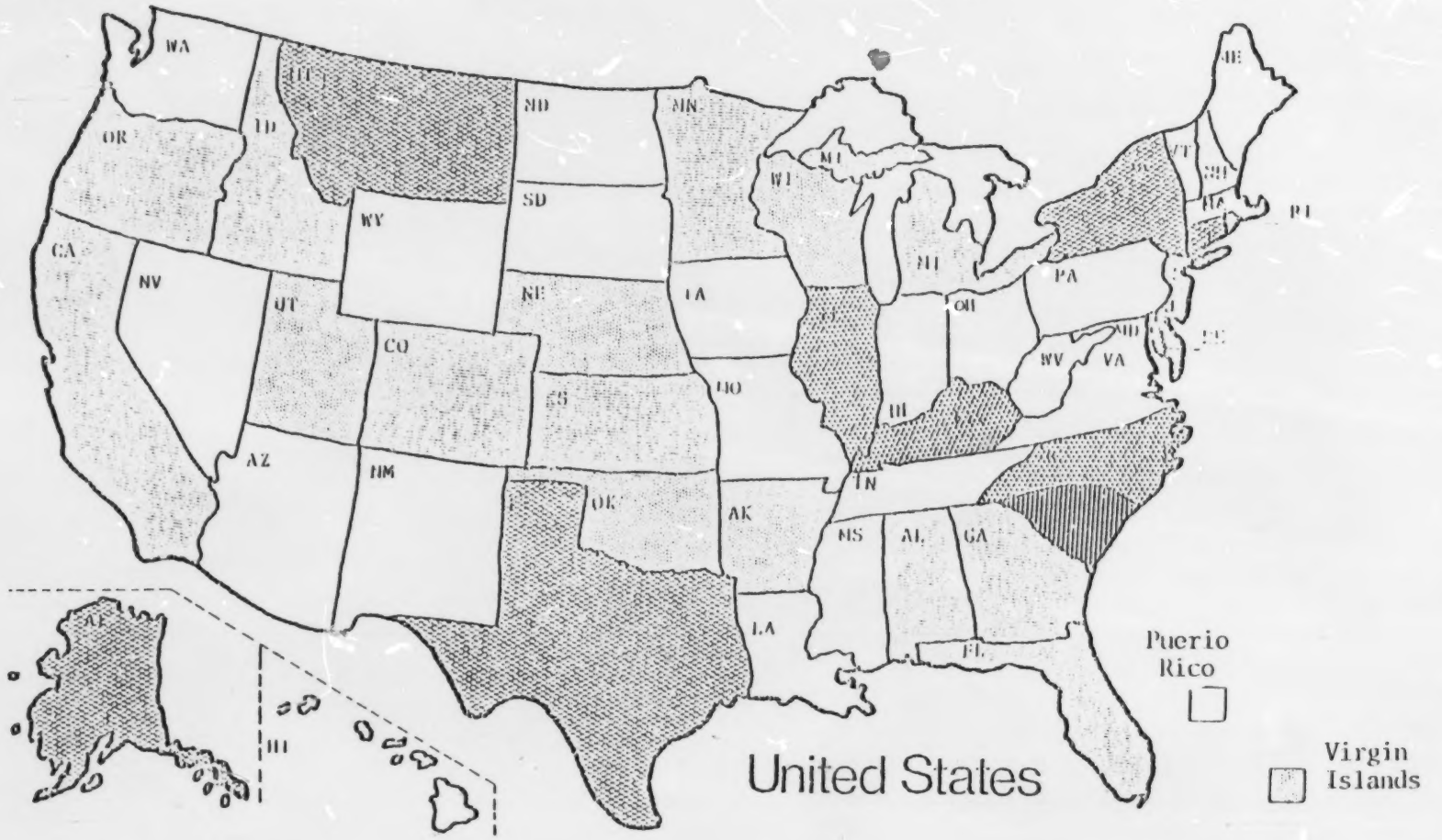

Figure 1. States Receiving Grants Through NIE's State Capacity Building Program (SCBP)

decisionmakers and practitioners can rationally consider alternatives to current practice and the results of research and development in improving educational programs." The design phase and first full year of the evaluation concentrate on describing the Program, clarifying the Program's objectives, and testing and refining instrumentation and data collection procedures; the second and additional years will concentrate more on delineating the impact of the Program in the states and on communicating evaluation results to NIE and the projects for their own improvement.

Two overarching questions focus all NTS evaluation activities:

- Is dissemination capacity being built? If so, how?
- Is the Capacity Building Program having an effect? If so, what is the nature of the effect?

The specific evaluation questions subsumed under these two overarching questions are listed in Appendix A.

The data collected by this effort will be used by several audiences. The National Institute of Education will use the data to aid in policy development and planning. The results of this study will also provide Congress and other decisionmakers with an evaluation of the effectiveness of the State Capacity Building Program in building capacity in educational dissemination. A goal of the evaluation is to develop a set of instruments and designs which could be used to monitor capacity building activities in all fifty-seven states and territories. States have indicated that the data also will be useful for their own internal evaluations and program planning. A separate technical assistance component provides tailored services for addressing project needs and will assist the projects in using evaluation information for internal improvement. In order to facilitate the use of the knowledge gained from this study, the evaluation also includes periodic meetings and reports to both NIE and the states.

Seventeen months of activities have culminated in the designs for the full-scale evaluation of the State Capacity Building Grants Program. The designs meet the two overriding objectives of the evaluation: (1) to provide NIE with evaluation information about the Grants Program and (2) to provide

states with programmatically useful information, as well as with limited technical assistance. In preparing these designs, NTS has viewed its task as part of the collaborative effort already underway between NIE and the states to develop nationwide capacity for improving education.

From the signing of the initial evaluation contract in September 1976 to the preparation of the final design documents in March 1978, a number of activities have been undertaken by NTS. The twelve major activities which have contributed to the final evaluation designs are depicted in Figure 2. All of these design activities employed qualitative methods. SEAs have been centrally involved in five design activities (i.e., familiarization site visits, policy implications analysis, instrument development, workshop series A and B, design review visits) and periphally involved in all others. While contributing to all aspects of the final designs, NIE has likewise been centrally involved in five of the major design activities (i.e., program announcements and evaluation RFP, NIE/NTS meetings, framework development, instrument development, policy implications analysis).

Six respondent groups have already been identified as participants in the first-year evaluation effort: State Capacity Building Program project directors, State Education Agency administrators, resource base administrators, linkage agents, clients of the State Capacity Building Program, and National Institute of Education program officers. In subsequent years of the evaluation, additional respondent groups (e.g., non-users) have been proposed. The respondent universe and sample size for the first year of the full-scale evaluation are summarized in Table 1. Figures pertaining to the universe of clients and agreed-upon sample size within each state are given in Table 2.

The proposed instrumentation for the first year of the evaluation was designed using a variety of qualitative methods including semi-structured interviews, informal discussions, brainstorming sessions, workshops, conferences, and policy implications analysis. Instrumentation for the first year of the full-scale evaluation includes both quantitative and qualitative

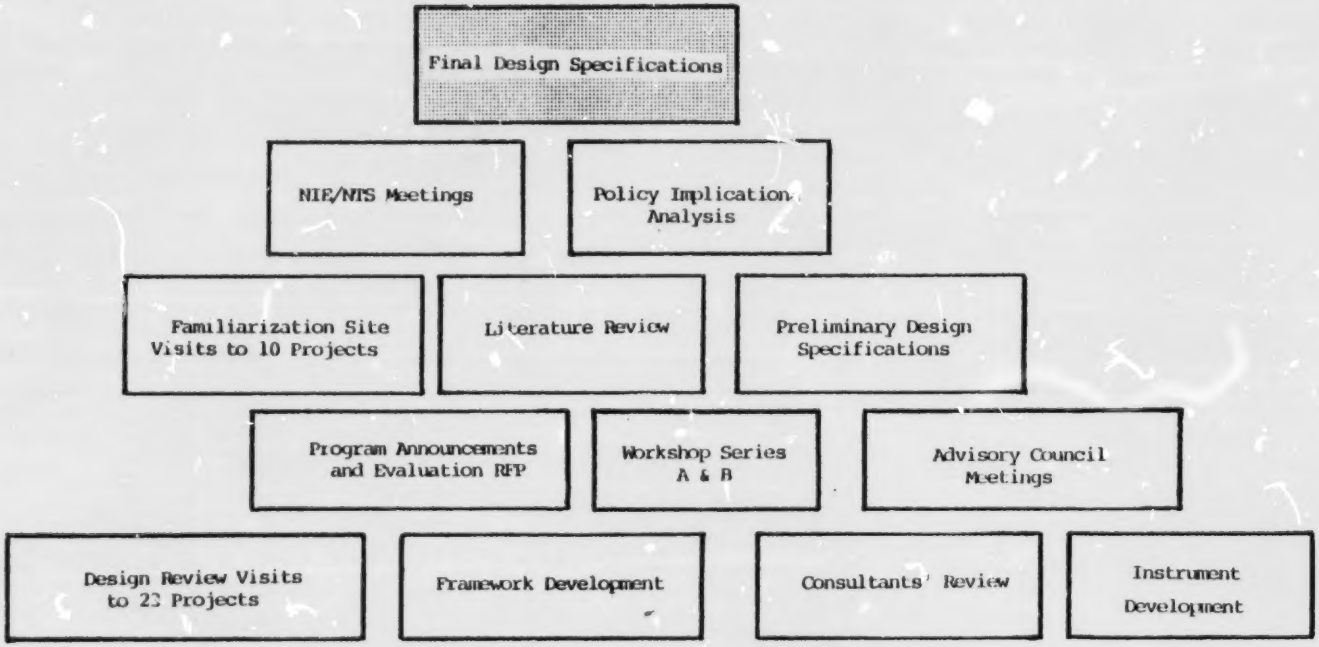

Figure 2. Twelve Activities Contributing to Final Design Specifications

Table 1. Respondent Universe and Estimated Sample Size for the First Year Evaluation

RESPONDENT	UNIVERSE SIZE	ESTIMATED SAMPLE SIZE
SCBP Project Directors	29	29
SEA Administrators	Unknown	145
Resource Base Administrators	75 - 100 (estimated)	25 - 35
Linkage Agents	7000 - 8000 (estimated)	70 - 90
SCBP Clients	6050 (estimated)	1730
NIE Program Staff	10	10

Table 2. Client Samples for States in the SCBP

STATE	UNIVERSE OF CLIENTS[1]	SAMPLE OF CLIENTS FOR EXTENDED PILOT[2]
Alabama	150	50
Alaska	400	100
California	150	50
Colorado[3]		
Connecticut[3]		
Delaware[3]		
Florida	35	35
Georgia	85	40
Idaho	550	150
Illinois	540	180
Kansas	1250	300[5]
Kentucky[3]		
Massachusetts[3]		
Michigan[4]		
Minnesota	200	100
Montana	700	100
New Hampshire[3]		
New Jersey[3]		
New York	2500	300[5]
Oregon	65	65
Rhode Island[3]	160	
S. Carolina	180	60
Texas	600	200

[1]These are upper bound estimates based upon three months for the extended pilot period.

[2]The number in this column represents the number of complete sets. Service Form, Process Form, Immediate Feedback Form and Client Assessment Form will be collected for each client in the sample.

[3]This state declined participation in the Client Assessment portion of the evaluation for this year.

[4]There are no clients per se as of this date in Michigan.

[5]These figures were stipulated by Kansas and New York.

tools. The following instrumentation has been proposed for the first year of the full-scale evaluation:

- Project Director Questionnaire
- Construct Definition Exercise
- SEA Questionnaire
- Project Director Interview Guide
- NIE Project Monitor Interview Guide
- NIE Program Officer Interview Guide
- Client Assessment Package
  - Service Form
  - Process Form
  - Linker Form
  - Client Immediate Feedback Form
  - Client Assessment Form

The three interview guides are qualitative instruments. The remaining instruments are comprised of structured, closed-ended items. Design activities planned in conjunction with the first full year of evaluation will result in a modified instrument package for the second year of the longitudinal evaluation.[7] First-year evaluation results also will be used to plan and develop subsequent technical assistance activities.

The collaborative spirit, which has been an explicit cornerstone of the State Capacity Building Program, provides the foundation for the NTS Evaluation. The collaborative interaction among information users and evaluators revealed a diverse set of information needs. Individuals (e.g., NIE program staff, state project staff, legislators, other federal program staff, and state, regional, and local education agencies) make decisions at different levels and require information in different formats at different times. A multi-method approach has been employed to clarify the information needs and to collect, analyze, and report information to alleviate identified needs.

III. BENEFITS ACCRUED IN INTEGRATING METHODS

The integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. In this section, the Evaluation of the State Capacity Building Program is used to illustrate the benefits obtained by the interplay of methods. Both the contributions of qualitative to quantitative methods and the reverse sequence are discussed. The methods are not always interchangeable.[8] Each method can, however, be strengthened by using intrinsic qualities of the other. The benefits to design, data collection, and analysis for both sequences are examined.

A. Contributions of Qualitative Methods to Quantitative Methods

Design

Qualitative methods can enrich quantitative designs by improving both the sampling framework and also the focus of the overall design. Exploratory interviews and observations can serve either to confirm or to rationalize a sampling framework (Sieber, 1974). Other qualitative tools can be used to pinpoint and prioritize information needs.

An evaluation can be conducted on either of two major units of analysis. One studies either <u>individuals</u> (e.g. clients) or <u>systems</u> (e.g. SEA dissemination systems, state dissemination systems) and asks what accounts for variation in their respective behaviors. In the NTS Evaluation, qualitative methods shaped the sampling framework at both the system and individual levels.

<u>The System Level</u>. Information obtained during the familiarization visits supported the original decision to collect identical data from all capacity building projects. Universe sampling at the project level appears to be the only adequate way to capture and describe the variations among states. Initial fieldwork also revealed that the inclusion of states without capacity building projects would provide additional information (in the form of comparison groups) about the NIE Program. Resource constraints, however, prohibited this fieldwork from dictating the sampling framework for the NTS Evaluation.

The dissemination system will be studied at three analytical levels: those of the project, the SEA, and the state. In the early years the longitudinal evaluation will focus on the project within the context of the SEA. In subsequent years, the focus will be expanded to study the state dissemination system in those states which have received capacity building grants.

The Individual Level. Based upon information obtained using qualitative methods, individuals were also included or excluded from the sampling framework. For example, during the familiarization visits, it became apparent that surveying all project-affiliated linkers would be a costly and inefficient method of obtaining information about linkage structures and processes within the state. In addition, such a survey could threaten a project's attempt to institutionalize existing linkages. Thus, in the final design a sample of linkage agents will be asked only to validate the project director's responses to questions referring to the organization and operation of the linkage system.

Qualitative methods were also used to focus the overall evaluation design. The information based evaluation (IBE) model was used to define and specify information requirements of the evaluation. Policy implications analysis (PIA) was designed by NTS to enliven the forum through which information users expressed their information needs. The construct definition exercise was developed to operationalize a definition for the concept "comprehensive state dissemination capacity."

Information Based Evaluation. As Alice learned from the Cheshire cat, if you do not know where you want to go, it does not matter in which direction you turn. It should come as no surprise that failure to involve the information users in the design phase drastically lowers the probability that the right questions will be posed and answered. To alleviate this problem, NTS developed and used the information-based evaluation (IBE) model (Stenner, 1974). The overriding consideration of this evaluation model is to supply information to individuals in decisionmaking roles. The focus of the IBE model has two closely related aspects: (1) the evaluation questions and (2) the ways in which the questions can be most usefully answered. The two reference points of this model are the information users and the information

needs as expressed in the evaluation questions. Using this model with NIE and the states, priorities were established in both information area categories (i.e., service capacity, user capacity, federal role) and the information users categories (i.e., NIE program staff, state project directors). The evaluation resources were then allocated to meet these identified priorities.

Policy Implications Analysis (PIA). Believing that it is the responsibility of the evaluation team to create a forum through which information users can express their information needs and feeling a need to enliven this forum beyond interviews, workshops, and conferences, NTS developed policy implications analysis (Stenner, A. J. and Madey, D. L., 1977). PIA is a program evaluation tool which draws upon several "futures" methodologies to account for the policy context surrounding an evaluation and to make more explicit the information requirements of key information users. PIA allows for the consideration of multiple possible contexts as information requirements are explicated. The technique extends and formalizes the IBE approach to delineating information domains and information users.

PIA launches policymakers into the future as they think about information needs. PIA is designed to elicit reactions to hypothesized evaluation findings to determine: (1) whether each finding fits within the expected scope of the evaluation; (2) whether each finding corresponds with expectations; (3) to what extent each finding has policy relevance; and (4) to what extent each finding is communicated in an appropriate way. PIA also asks for policy actions that might result from the findings, what level or type of information would be needed to justify the findings, and whether additional information would be needed to justify the policy actions.

First used in the evaluation of the State Capacity Building Program, policy implications analysis provided a window through which information requirements of NIE and state project personnel became clearer. Results were used to guide the evaluation team in projecting decision problems and associated information requirements. After the evaluation questions and requirements had been finalized, the evaluation team was able to proceed more systematically with instrument development.

Construct Definition Exercise. NIE's decision not to be prescriptive meant that the Institute elected not to define capacity building in measurable, operational terms. To do so would have been to intervene in the 29 "experiments" which embody the various ways in which capacity is being built. The non-prescriptive stance precluded NIE from providing an operational definition of "comprehensive state dissemination capacity." Rather, NTS worked with the states to develop a provisional operational definition. The technique developed by NTS was termed the construct definition exercise. The evaluation team in collaboration with the states generated six primary constructs that were viewed as necessary and sufficient to specify the meaning of "comprehensive state dissemination capacity." Each construct was defined in terms of observables, or more precisely indicants, which, when taken singularly, provide only partial explication, but when taken collectively, fully define the constructs. Through this exercise, NTS in collaboration with the states forged an operational definition for NIE's concept of "comprehensive state dissemination capacity."

Data Collection

Observations and exploratory interviews which precede quantitative survey instruments can supply useful information about the receptivity, frames of reference, and attention spans of respondents (Sieber, 1974). When systematic pretesting is not undertaken because of time or financial constraints, it is still possible to obtain valuable information by using qualitative methods prior to the full-scale data collection phase.

Qualitative pretesting serves two purposes. First, it results in a better instrument package, an objective of all design work. Second, the rapport which stems from qualitative pretesting smooths the way for more elaborate, time-consuming, and sometimes threatening aspects of full-scale data collection.

In the NTS Evaluation, informal pretesting of the proposed Client Assessment Package (CAP) during the design review visits and spring workshop series served both purposes. First, qualitative pretesting resulted in significant improvements in the original instrument package. Certain items (e.g., client's race) were eliminated and others were refined (e.g., follow-up assistance) in order to reduce respondent burden and eliminate potentially offensive items. In addition, several previously excluded topics (e.g.,

satisfaction with personal services provided) were identified as important aspects of client assessment and the instrument package was modified accordingly. Second, the informal pretesting established a rapport between the individual project staffs and the evaluation team, which should smooth the way for the actual implementation of the instrument package. The projects early recognized the value of the CAP in terms of information yield and eagerly await this program improvement tool. The contacts, approval, and support of high level administrators within the SEAs and information resource bases should facilitate subsequent contacts with actual clients.

Data Analysis

Information gathered using qualitative methods can aid in the analysis and interpretation of quantitative data in at least six ways (Sieber, 1974). First, the conceptual framework that guides the analysis can be derived, in part or in whole, from qualitative work. Second, in areas where methods yield information overlap, certain quantitative results can be verified by results obtained using qualitative methods. Third, qualitative data can be used as the basis for selecting survey items to be used in index construction. Fourth, external validation of empirically-generated constructs can be obtained by comparison with observation data. Fifth, case studies that illustrate statistically derived models can be supplemented by field observations. Sixth, clarification of ambiguous and provocative replies to questionnaires can be obtained by reexamining field notes. Examples from the Evaluation of the State Capacity Building Program and other studies illustrate the six ways in which qualitative methods can illuminate the analysis of quantitative data.

Conceptual Framework Development. In order to conduct a thorough and comprehensive evaluation, there must be an organizing framework which builds upon the information needs, relevant theory, and knowledge of the program. A conceptual framework is a necessary step in evaluation design (Stenner, 1974; Scriven, 1974; Stufflebeam, 1974). Its function is to act as a data organizer; it guides the data analysis.

The conceptual framework that guides the current evaluation is a product of seventeen months of design activities. Six information sources, all qualitative, provided heuristic bases for framework development. First, a

considerable amount of descriptive information was collected during the design phase, particularly from the first cohort of states.[9] Second, additional state visits clarified staff perceptions concerning the perceived differences and similarities among projects.[10] Third, completion of an expanded literature review characterized the state of the art in dissemination, knowledge production and utilization, organizations, and implementation theory.[11] Fourth, NTS conducted three workshops during which state personnel reviewed the designs, instrumentation, and data collection plans. Fifth, experts in the areas of dissemination and implementation were used as consultants and contributed to the development of the organizing framework.[12] Finally, two draft reports were reviewed by the National Institute of Education, and suggestions for changes were incorporated in the final conceptual framework.[13] Each of these qualitative methods contributed to the conceptual framework, which undergirds the structured data analysis plan.

Verification. It is sometimes possible to validate (or in some cases invalidate) quantitative findings by referring to information obtained using qualitative probes and interviews. Since the NTS Evaluation has not yet completed a full-scale data collection effort, actual illustrations from the current study are presently nonexistent. The examples described here are from a previously completed study of educational research organizations, (Sieber and Lazarsfeld, 1966) in which the authors were able to verify survey findings by reference to subsequent fieldwork:

> Tabulation of the questionnaire showed that extremely few doctoral recipients who had worked in research bureaus as assistants remained as staff members... It was felt that this survey finding might explain the lack of continuity in research bureaus, the difficulty of recruitment, and the strong influence of each succeeding director... But since only about two-thirds of the respondents had answered this question, the researchers felt uneasy about resting their case on the survey finding alone. During later informal interviewing, the directors were asked how they felt about retaining research assistants as professional staff members. With only one exception, the dozen or so directors believed that students should be encouraged to leave the bureau after getting their degrees. The reason given was that students would not become independent of their mentors unless they took positions elsewhere. Since this viewpoint was expressed with great conviction by the informants, the field interviews lent plausibility to the survey finding.[14]

The invalidation of survey results by qualitative methods has also been reported:

> In her study of working-class marriage, it was very important for Komarovsky (1962) to classify her subjects according to differing degrees of marital happiness in as reliable a manner as possible, for marital happiness was a crucial dependent variable. She therefore drew upon information gathered in a series of detailed and indirect probes. Comparing her distribution of cases with large, representative samples of the same social strata which employed more direct self-ratings, she found that her own population contained a larger proportion of unhappy marriages... Komarovsky accounts for this discrepancy by reference to the more subtle techniques of qualitative case study, making it more difficult for the respondent to conceal the unpleasant aspects of marital relations... Consequently, instead of being misled by the results of typical survey items, Komarovsky employed a more qualitative approach when classifying her subjects according to certain major variables in her study.[15]

Qualitative verification of a quantitative survey instrument usually focuses on selected items or variables, as in the above examples. It would be possible, albeit tedious, to extend such verification to an entire instrument.

Index Construction. Indices for use in analysis of quantitative data may be constructed from qualitative observations, probes, and interviews. In the NTS Evaluation, scales for the construct definition instrument were derived from discussion with project directors during the workshops. For example, the variables used to define the construct "comprehensiveness of linkages" include availability (yes or no) and degree of use (five-point scale ranging from very extensive to very limited) while the procurement index for comprehensive resource base includes four points (refers, provides directly, serves as a broker, does not provide).

External Validation. A statistical index may be validated by reference to qualitative work. Although actual examples have not yet occurred in the NTS Study, hypothetical examples do come to mind. Quantitative measures obtained using the construct definition exercise could be tested by using qualitative methods. For example, the measure of "institutionalization" could be verified by using

interviews with project directors, high level SEA administrators, directors of other state dissemination efforts, and selected clients. During interviews, an attempt would be made to explore the amount of institutionalization which has occurred. Individuals might even be told where their project had been ranked on the construct and then asked if the score coincided with their perceptions. If discrepancies were discovered, additional probing would be used to determine whether the mismatch were due to a weakness in the instrument or misinterpretation by the individuals interviewed.

Case Studies. The use of case studies to illustrate models or statistical and historical types is a common practice in the social research field (Kahl, 1953; Gordon, 1957; Komarovsky, 1962; Sieber and Lazarsfeld, 1966). It has also been proposed in recent evaluations. For example, case studies developed by Education Turnkey Associates to describe the actual implementation of P.L. 94-142 will be used to illustrate the implementation models derived in SRI's longitudinal study of progress in implementing the same legislation (Stearns, M. and David, J.L., 1977).

In the Evaluation of the State Capacity Building Program, models of capacity building will be derived using quantitative methods. To illustrate the models, case studies will be used. Either the case studies will be obtained from those which then exist or else case studies will be developed specifically by NTS to illuminate the models.

Clarification. A final contribution of qualitative work to analysis of quantitative data involves the clarification of puzzling but provocative replies to written questionnaires. Since examples of this final contribution do not presently exist in the NTS Evaluation, an illustration has been excerpted from another study:

> In a survey of the directors of educational research bureaus, the following question was asked: "In general how fruitful have interchanges been with academic departments in the university; what problems have been encountered, if any; and what directions would you like future interchanges to take?" One director wrote the following reply: "Professors in the liberal arts seem not to be able to make advancements within their respective departments if they participate heavily in the activities

> of the Center." The response was curious, possibly significant, but far from clear. Later, in the course of fieldwork among selected bureaus, the director was asked to clarify his answer. He explained that academic personnel who became associated with his organization lost visibility in their departments. Their frequent absence from the department was interpreted as a lack of departmental commitment. His clarification illuminated the problem of integrating research bureaus into the universities, which became a dominant theme in subsequent thinking.[16]

A similar integration of techniques is anticipated in the ongoing study. Specific portions of workshops with state project directors will be devoted to clarifying quantitative data puzzles.

B. Contributions of Quantitative Methods to Qualitative Methods

Quantitative techniques may contribute to qualitative methods in many ways. The contributions of such techniques as surveys and questionnaires to interviews and case studies are perhaps less appreciated than the reverse. In some instances it would seem to be "methodologically obligatory" for qualitative methods to take advantage of quantitative techniques (Sieber, 1974). As in the previous section, illustrations are derived, whenever possible, from the NTS Evaluation.

Design

Quantitative methods are useful for identifying the most suitable systems or individuals to be studied using qualitative methods. Either a representative or unrepresentative system or individual may be identified using quantitative statistical methods.

The System Level. One anticipated outcome of the efforts of the early years of this evaluation is the ability to classify projects into types or models which describes the states' approaches to capacity building, or, in other words, the states' dissemination systems. The models will have been derived, for the most part, using quantitative methods. Clustering, a step which precedes model building, will be done quantitatively using nominal or ordinal algorithms. Once models are developed, the evaluators will be able to

select projects that represent different approaches to capacity building for more intensive study. Similarly, if early evaluation results reveal that a certain project does not fit any model, it could be selected for more in-depth study. Such outlier studies can reveal model imperfections, an understanding of which can lead to model or framework revisions.

The Individual Level. The Client Assessment Package is comprised of five instruments with which client satisfaction and information utilization will be studied. Quantitative methods (e.g., path analysis, cluster analysis, contingency table analysis) will be used to analyze the client data. Based on results obtained, case studies of users and non-users are planned for subsequent years of the evaluation. One state with a low (16%) return rate on feedback forms has already suggested an in-depth study of the people who don't respond to such instruments. Others are interested in detailed studies of repeat users.

To summarize, quantitative methods can improve the design of a qualitative study by identifying both representative and unrepresentative cases at either the system or the individual level. Studying representative cases serves the goal of generalizability; studying unrepresentative cases facilitates refinement of models and theories.

Data Collection

Three contributions of quantitative work to the collection of qualitative data are possible. First, quantitative data can provide leads for later interviews and observations. Second, quantitative data can provide information about informants or subjects who were overlooked. And third, quantitative data can correct for the "elite bias" in the interpretation of events (Sieber, 1974).

Leads. When quantitative work has been conducted prior to qualitative work, the replies to survey instruments can provide leads for later interviews and observations. The need to ask routine background questions during an interview is eliminated when such questions have been included in a previously administered survey. When data collection for the in-depth user and non-user studies begins in the NTS Evaluation, guidelines and probes for formal interviews

(phone and/or personal) will be derived from items obtained using the Client Assessment Package. The use of previously obtained information will facilitate qualitative data collection in two ways: first, it will eliminate the need to ask routine background questions of the clients selected for case study; second, it will establish a bond between interviewer and interviewee, thus increasing the rapport between the two. Similarly, data collected using the construct definition exercise may result in guidelines and probes for interviews in subsequent years of the evaluation. The quantitative data may also feed into subsequent iterations of the PIA exercise.

Overlooked Respondents. Quantitative data can also provide information about informants or subjects who were overlooked. As a hypothetical illustration, if the data collected using the Client Assessment Package consistently reveals that the most important person in facilitating use of an information service is the local librarian, and if the NTS Study has not examined the role of the local librarian at the time of such a discovery, then such a study would be required to understand further the domain of user capacity.

Correction of the Elite Bias. One pitfall in qualitative data collection is an "elite bias" in the selection of informants and in the evaluation of statements (Sieber, 1974). If an evaluation gravitates to the elite of a social system in the course of the qualitative work, the results may be biased. Studying only the elite is a distinct possibility in an evaluation for four reasons. First, the elite or "gatekeepers" are often an evaluator's initial contact with a group. Since first impressions are often lasting ones, the views of the elite may color the evaluator's objectivity throughout the ensuing study. Second, if the elite individuals are respected in the larger social system, the evaluator may neglect association with other contacts. Third, the elite individuals may be more articulate than other members of the group, and thus appear to be more knowledgeable. Fourth, an evaluator who spends more time collecting information from the members of the elite may give more weight to their viewpoints than to those of other group members.

The utility of integrating methods to correct the elite bias is illustrated by using a study that initially set out to examine the structure of two suburban school systems and developed into a study of school boards, superintendents, and leaders of high school teachers:

> Prior to looking at the results of the survey, I predicted the proportion of teachers who would respond in particular ways to survey questions. I then compared my predictions with the actual responses. It became obvious when observing these comparisons that I had unwittingly adopted the elites' version of reality. For example, I overestimated the extent to which teachers felt that the administration accepted criticism... Similarly, I had assumed that the teachers were more satisfied with evaluative procedures than was in fact the case... I also overestimated the rank-and-file support for the leaders of the teachers association, with whom I had spent a good deal of time. In short, I had fallen prey to the elite bias, despite recent training in the dangers of giving greater weight to prestigious figures as informants. The survey not only constrained me to see that my qualitative data-collection procedures had been faulty, but also provided the opportunity to learn about an entire stratum which I was now aware of having glossed over in the fieldwork, namely, the elementary teachers. Apparently, the elite bias had operated also in my preference for secondary teachers, who are the more esteemed both in the profession and the community.[17]

Data Analysis

Quantitative methods can contribute to the understanding of qualitative observations in at least three ways: (1) by correcting the holistic fallacy, (2) by verifying qualitative interpretations, and (3) by casting a new light on field observations (Sieber, 1974). Of all the synergistic effects gained by the interplay of methods, these contributions are perhaps most overlooked.

Correction of the Holistic Fallacy. The "holistic fallacy" is the "tendency on the part of field observer to perceive all aspects of a ... situation as congruent."[18] This fallacy is a possibility in the Evaluation of the State Capacity Building Program. An articulate state project director who feels that he/she is doing an outstanding job of developing a

computerized resource base may give the impression that the other required components (e.g., linkage, leadership/management, client base) are developing well, when in fact quantitative methods (e.g., the Client Assessment Package) might show that few clients actually have access to the information resource base and even fewer are capable of using the information in the form provided.

Verification of Qualitative Interpretations. Verifying observations collected during informal visits is possible using quantitative instruments. In the NTS Study, hypotheses crystalized during the design visits will be explored more systemmatically during full-scale data collection. For example, it appears that the presence of a regional service arrangement influences the overall structures and processes of service capacity. Similarly, it seems as if the strong support of a chief state school officer is a prerequisite to institutionalization of dissemination functions within an SEA.

Casting New Light on Qualitative Observations. Quantitative findings can sometimes illuminate a field observation that was inexplicable or misinterpreted. Serendipity is not limited to qualitative work. The analysis of quantitative data often yields unanticipated observations and concepts. It has been argued that surveys are uniquely qualified for the measurement of unanticipated concepts (Schenkel and Sieber, 1969). Since NTS has not yet completed a full-scale data collection phase, an illustration from another study is presented:

> In a study of two surburban districts, it was observed that a smaller proportion of teachers turned out to vote in the bond issue election in the larger district. When this observation was shared with informants, many explanations were offered. The poor turnout had been tentatively attributed to the alienation of many teachers in the more bureaucratized system... While perusing the distribution of responses to a survey, the investigators noticed with surprise that 39% of the teachers in the larger district resided outside of the district, compared with only 18% in the smaller district. The teachers in the larger district were simply less often legally qualified to vote. The observations of poorer turnout was therefore reinterpreted.[19]

In the NTS Evaluation, similar unanticipated findings may emerge from the full-scale data collection effort. Such findings will be shared with NIE personnel and state project directors in an effort to reduce alternative hypotheses and more accurately to explain capacity building--its structures, processes, and outputs.

In summary, benefits are gained by employing both quantitative and qualitative methods in one study. "The methods are analogous to zooming in and zooming out with a lens. To the extent that they are reproduced objectively, wide-angle, telephoto, and microscopic views must be simultaneously valid, and zooming from different directions merely focuses attention on different facets of the same phenomenon... There are no grounds, logical or otherwise, for calling any view simple. We can start anywhere and zoom in to infinite detail, or zoom out to infinite scope."[20]

Qualitative methods provide clues to some of the causes behind the changes (or lack thereof) observed using more quantitative methods. Qualitative methods sharpen the focus on and provide additional explanations of the observed relationships. Quantitative methods yield objective instruments which can be used repetitively with little additional expenditure of resources. Quantitative methods provide a lasting, enduring yardstick for measuring change.

The examples presented in this section underscore that timing considerations are very important to the management of an evaluation which seeks to benefit from both methodologies. Three time sequences are possible:

(1) Qualitative methods can precede quantitative methods.
(2) Qualitative and quantitative methods can proceed concurrently.
(3) Qualitative methods can follow quantitative methods.

Each sequence is most prevalent in a separate phase of the evaluation. The first sequence is most frequent in the design phase; the second sequence, involving a concurrent use of methodologies, is most often associated with the data collection phase; and the third sequence is prevalent in analysis.

In the NTS Evaluation, qualitative methods utilized during the design phase contributed to the formulation of a conceptual framework and the development of more quantitative instruments (e.g., Client Assessment Package). During the data collection phase, concurrent use of both

methodologies will augment the information yield. Since the evaluation team will be in the field during data collection, it will be possible to learn something about the actual meaning of the questionnaires to respondents. To some extent, it will be possible to assess the obtrusiveness of the quantitative instruments and take such qualitative judgments into account during interpretation of the data. In the data analysis phase, findings obtained using the quantitative instruments will be clarified and extended using, once again, qualitative methods (e.g., workshops, "data-massaging" conferences). The information users will be involved in the transformation of data into findings, conclusions, and recommendations.

By interweaving qualitative and quantitative methods over the duration of the evaluation, regardless of the primary method during any phase, the utility of the information provided to decisionmakers at the federal and state levels should be increased. Integrating methods in a single evaluation yields a richer tapestry of findings and a deeper understanding of the program--its structures, processes, and outcomes. Decisionmakers are more likely to understand, value, and use evaluation results when they have been intimately involved in the evaluation.

IV. COSTS INVOLVED IN INTEGRATING METHODS

Perhaps no evaluation is without potential disadvantages, and one which integrates methods is no exception. While integrating methods does alleviate the criticisms of a single method approach (cf. Webb, 1966; Mason, 1973), the evaluator encounters a new set of problems when quantitative and qualitative methods are integrated within a single evaluation. The major disadvantages are increased costs in time, resources, and personnel.

Time

A longer time period is needed to design, implement, and successfully complete an evaluation which integrates methods. An optimal schedule for this type of evaluation involves a period of qualitative fieldwork prior to the development of more quantitative instruments. The sharing or collaborative process involves a great expenditure of time. Just reviewing

the documents and instruments with NIE and the states involved two weeks of regional workshops and several NIE meetings extending over a four month period. Preparing instruments in committee was a tedious and sometimes frustrating process. In addition, interviewing involves more time than administering quantitative instruments and places greater burdens on respondents (thus the Office of Management and Budget's increased scrutiny of such instruments).

The NTS Study completed a seventeen-month design phase before beginning the full-scale evaluation. The extended design phase has caused legitimate concerns among the potential information users. Since evaluation information from this study is not yet available, program improvement and technical assistance efforts must be based on other sources of information.

Resources

Closely correlated with a greater expenditure of time is the greater expenditure of resources. The most obvious costs incurred in an evaluation which extends over a lengthy time period are the salaries which must be paid for the study's lengthier duration. But greater expenditures for salaries are not the only costs incurred.

More demands are placed on all information users. Their presence is necessary at design conferences and workshops. Travel costs, personnel costs, and subsistence costs are encountered in any workshop gathering. In the NTS Evaluation, state project directors traveled literally from all corners of the United States to participate in design conferences. A series of regional workshops, rather than one centrally located conference, was planned in order to decrease travel costs for participants. Thus, Alaska's project director traveled to the San Francisco workshop; New Hampshire's project director attended the Philadelphia workshop; and Florida's project director participated in the Atlanta workshop. While easing the burden for project directors, regional workshops increased the travel burden on the NTS evaluation team.

The sharing or collaborative process which is an integral part of an evaluation has its own associated costs. The mere printing and binding of

enough copies of the two-volume design document to mail to states cost almost $700.00. Preparing five copies for NIE would have cost much less.

The integration of methods necessitates certain modifications of traditional evaluation designs; these modifications also involve a greater expenditure of resources. Adapting the typical large-scale survey of individuals so as to profit from qualitative work entails additional costs. If informal networks can be identified among the individuals, they can be studied in much the same way that an evaluation deals with a more formal system. If this type of design is neither feasible nor sensible, then every *n*th interviewer could be instructed to make certain observations or extend the survey into an unstructured interview format. Such interviewers would either have to receive special training in interviewing or else have previous experience in such fieldwork. The costs involved in integrating methods are reduced when studying a small number of formally organized systems, such as schools, since the respondents are clustered within settings which have definite boundaries.

Modifications are also necessary if a traditional qualitative design is to take advantage of quantitative methods. In conjunction with planned case studies, certain clusters or networks of respondents might be identified and then selected individuals could receive questionnaires in order to provide statistical information. Such modifications also involve an additional expenditure of resources.

Personnel

The optimal staff for an evaluation which integrates methods would be comprised only of individuals with expertise in both methodologies. An alternative is to use an evaluation team in which the different members specialize in one type of methodology but can also acquire skills in the other. Whenever possible, and especially when a lengthy design phase is not practical, individuals who possess a rich knowledge of the field under investigation should be employed.

Adequate time, resources, and personnel must be available to meet the increased costs incurred in an evaluation which integrates methods. Otherwise, undertaking such an evaluation would be futile.

V. APPLICABILITY TO OTHER EVALUATIONS

The adjustments in traditional evaluation designs necessitated by the integration of quantitative and qualitative methods appear to produce another evaluation style. In terms of methods, one plus one equals three. And what's the "three?" The interplay between the two types of methods. The interaction, the synergistic coming together which creates something that never existed before.

The integration of quantitative and qualitative methods within a single evaluation has synergistic effects in the three major phases of design, data collection, and analysis. The methods are not always interchangeable; each method can, however, be strengthened by using intrinsic qualities of the other.

To benefit from both methodologies, an evaluation must be able to incur increased costs in time, personnel, and resources. When these costs are not exorbitant, integrating methods will increase and enhance the information yield and, by so doing, facilitate the process of clarifying and alleviating information needs. A multi-method approach is well suited to meeting the diverse needs of all the information users involved in a collaborative evaluation.

In a study which seeks to know both whether a program works and also why it works (or does not work), integrating methods offers enormous potential for improving program evaluations. In the State Capacity Building Program Evaluation, both types of information were needed and an evaluation which integrated methods was designed to meet the information requirements.

At present, examples of this "integrated evaluation" are too scarce to deduce all the general principles to be followed in organizing future evaluations. As more studies undertake an integrated approach to evaluation,

these principles will emerge. In the meantime, evaluators should be aware of all available evaluation methods and should choose their methods in terms of their value for producing the kinds of information needed in any given study, rather than in terms of some preconceived biases about what constitutes legitimate "evaluation methods." In making these choices, evaluators should examine the full spectrum of available evaluation methods. When existing methods are inappropriate, evaluators should attempt to develop, implement, and test new methods. When selecting methods and designing a study, evaluators should be receptive to the enormous potential which integrating methods offers for improving evaluations.

FOOTNOTES

1. The classical experimental design has also been designated the "agricultural-botany" design. This terminology is not accidental. Many of the statistical and experimental techniques utilized in educational evaluation were originally developed (e.g. by Fisher) for use in agricultural experiments. Parlett and Hamilton (1975) concisely describe the most common form of agricultural-botany type evaluation: [It] "involves the assessment of the effectiveness of an innovation by examining whether or not it has reached required standards on prespecified criteria. Students--rather like plant crops--are given pretests (the seedlings are weighed or measured) and then submitted to different experiences (treatment conditions). Subsequently, after a period of time, their attainment (growth or yield) is measured to indicate the relative efficiency of the methods (fertilizers) used. Studies of this kind are designed to yield data of one particular type, i.e., "objective" numerical data that permit statistical analyses. Isolated variables like IQ, social class, test scores, personality profiles, and attitude ratings are codified and processed to indicate the efficiency of new curricula, media, or methods." (p. 142)

2. Peter H. Rossi and Sonia R. Wright, "Evaluation Research: An Assessment of Theory, Practice, and Politics," Evaluation Quarterly, 1 (February 1977), 13.

3. President Richard M. Nixon, Economic Opportunity Message to the Congress, February 19, 1969, as quoted in Williams, Walter and Evans, John W., "The Politics of Evaluation: The Case for Head Start," The Annals of the American Academy of Political and Social Science, 385 (September 1969), 118.

4. Robert L. Wolf and Barbara Tymitz, "Toward More Natural Inquiry in Education," n.d. (mimeo)

5. Interestingly, a similar polemic existed in the field of sociological research between the proponents of qualitative fieldwork and the advocates of survey research. In contrast to the evaluation field, the early preeminent methodology was qualitative fieldwork (as exemplified by such classics as the Hawthorne studies, the Middletown volumes, and the Chicago studies of deviant groups). Following World War II, survey research became the dominant methodology and fueled a debate between the proponents of older field methods and advocates of newer survey techniques. The debate in the social research field was first mediated by the assertions of Trow (1957) and Zelditch (1962) that the nature of the problem dictates the method to be applied. In discussing the broader field of research in the social sciences, Webb et al. (1966) suggested the use of multiple methods or "triangulation" as an alternative to the single method approach to research. Sieber (1973), stimulated by the

ideas of John D. Ferguson, was perhaps the first to document the benefits to be gained by the integration of fieldwork and survey methods in social research. Sieber's arguments, although overlooked by evaluators, are as applicable to the evaluation field as they are to the field of social research. I am especially indebted to Sieber for the ideas presented in this paper regarding the interplay of methods and for many of the examples presented in Section II.

6. The polemic has also obscured the fact that each camp actually employs, although not necessarily documents, methods of the other end of the continuum. Another way to reconceptualize what appears to be a false dichotomy is to expand Guba's wave conceptualization of naturalistic inquiry (1978). According to this conceptualization, the naturalistic investigator cycles through a series of observations that are directed first at discovery and then at verification. Through the wave of discovery and verification, the investigator expands and reduces the domain of inquiry using a mixture of naturalistic strategies. The wave conceptualization can be expanded to include qualitative and quantitative methods and reshaped into a spiral (or series of decreasing waves) to represent closure or convergence. (Guba does discuss convergence in his monograph, but does not incorporate this concept in his wave theory.) In the spiral conceptualization, qualitative and quantitative methods interact synergistically throughout the course of the investigation or until closure is reached.

7. For a more extensive description of the proposed evaluation, readers are referred to the following documents: Madey, D.L.; Haenn, J.F.; Everett, M.; and Strang, E.W.; A Conceptual Framework for the Evaluation of the State Capacity Building Program: Durham, N.C.: NTS Research Corp., March 1978 (in press) and Haenn, J.F.; Madey, D.L.; Everett, M.; and Strang, E.W.; Request for OMB Clearance with Supporting Documents for the Evaluation of the State Capacity Building Program. Durham, N.C.: NTS Research Corp., April 1978 (in press).

8. Sieber (1972) is not entirely correct in his interpretation (p. 1340) of Webb's (1966) argument in behalf of multitechniques. Webb does not argue that the methods are entirely interchangeable as Sieber states. Webb is careful to add that using multiple techniques could "supplement" as well as "cross-validate" a specific technique (p. 2). While Sieber's interpretation of Webb's argument is inaccurate, his own thesis is complete and fairly concise: "there are areas of informational overlap between methods, but there are also large areas of information which can be gained only by a particular technique. If each technique has an inherent weakness it also has an inherent strength unmatched by other techniques. The opinions held by a large population can be measured only by survey techniques; the unverbalized normative pattern of a small group might be measureable only by observation." (p. 1337)

9. Two documents prepared during the design phase provide substantial documentation about the individual projects comprising the State Capacity Building Program. 1977 State Abstracts (Madey, D.L.; Mojkowski, C.; and Strang, E.W.) contains summaries of twenty-four capacity building

projects. Included are the ten SEAs initially funded in 1975 and the fourteen additional projects funded in 1976. State Reports (Madey, D.L.; Mojkowski, C.; and Strang, E.W.) contains more extensive documentation on nine of the first states funded through NIE's Capacity Building Program.

10. Chapter II of State Reports (Madey, D.L.; Mojkowski, C.; and Strang, E.W.) documents perceived differences and similarities among the nine projects remaining in Cohort I.

11. Building Capacity in Dissemination: Literature Review, (Strang, E.W.; 1977) is a general summary of dissemination literature. During the refinement and expansion of the conceptual framework, a broader spectrum of literature was reviewed and incorporated into The Conceptual Framework.

12. Individuals consulted during the development of the conceptual framework include Dr. Michael Fullan, Ontario Institute for Studies and Dr. Karen S. Louis, Abt Associates. NIE colleagues reviewing the framework include Ward S. Mason, Mary Ann Millsap, Charles Haughey, John Coulson, John Egermeier, Carnot Nelson, and Naida Bagenstos. NTS Advisory Council members consulted during the design phase include Mr. Richard Brickley, Dr. David Clark, Dr. Paul Hood, and Dr. Sam Sieber.

13. D.L. Madey, J.F. Haenn, M. Everett, and F.W. Strang, A Conceptual Framework for the Evaluation of the State Capacity Building Program, Durham, N.C.: NTS Research Corp., March 1978 (in press).

14. The Sieber and Lazarsfeld study is described in Sieber, Sam D., "The Integration of Fieldwork and Survey Methods," American Journal of Sociology, May 1973. 28:6, p. 1346.

15. The description of Komarovsky's study was excerpted from Sieber, Sam D., Ibid., pp. 1346-1347.

16. The description of Sieber's survey of educational research bureaus was reported in Sieber, Sam D., Ibid., p. 1350.

17. Sieber, Sam D., Ibid., p. 1353.

18. Sieber, Sam D., Ibid., p. 1354.

19. Sieber, Sam D., Ibid., p. 1357.

20. Edwin P. Willems and Harold L. Raush, Naturalistic Viewpoints in Psychological Research, New York: Holt, Rinehart, and Winston, 1969, pp. 82-83.

REFERENCES

Binder, A. "Statistical Theory." In P. R. Farnsworth, O. McNemar, and Q. McNemar (Eds.), Annual Review of Psychology, 1964, 15: 277-310.

Boring, E. G. "The Nature and History of Experimental Control," American Journal of Psychology, 1954, 67: 573-589.

Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Chicago, Illinois: Rand McNally, 1966. Originally published in Gage, N.L., Ed., Handbook on Research on Teaching, Chicago, Illinois: Rand McNally, 1963, 171-246.

Cooley, William W. and Lohnes, Paul R. Evaluation Research in Education: Theory, Principles, and Practice. New York: Irvington Publishers, Inc., 1976.

Douglas, Jack D. Investigate Social Research. Sage Library of Social Research, No. 29, California: Sage Publications, Inc., 1976.

Evans, John. "Evaluating Social Action Programs," Social Science Quarterly, 1969, 50:3.

Fisher, R. A. The Design of Experiments. Edenburgh: Oliver and Boyd, 1935 et.seq.

Gordon, C. Wayne. The Social System of the High School, Glencoe, Illinois: Free Press, 1957.

Kahl, Joseph A. "Educational and Occupational Aspirations of 'Common Man' Boys," Harvard Educational Review, 1953,23 (Summer): 186-203.

Guba, Egon G. Toward A Methodology of Naturalistic Inquiry in Educational Evaluation. Indiana University: School of Education, 1978 (in press).

Haenn, Joseph F.; Madey, Doren L.; Everett, Michael; and Strang, Ernest W. Request for OMB Clearance with Supporting Documents for the Evaluation of the State Capacity Building Program in Dissemination. Durham, N. C.: NTS Research Corporation, April 1978 (in press).

Herriott, R. E. and Muse, D. N. "Methodological Issues in The Study of School Effects," in Fred N. Kerlinger (Ed.), Review of Research in Education], Itasca, Illinois: F. E. Peacock, 1973: 209-236.

Komarovsky, Mirra. Blue-Collar Marriage. New York: Random House, 1962.

Madey, Doren L.; Haenn, Joseph F., Everett, Michael; and Strang, Ernest W., A Conceptual Framework for the Evaluation of the State Capacity Building Program. Durham, N.C.: NTS Research Corp., March 1978 (in press).

Madey, Doren L.; Mojkowski, Charles; and Strang, Ernest W., 1977 State Abstracts. Durham, N.C.: NTS Research Corp., September 1977.

Madey, Doren, L.: Mojkowski, Charles; and Strang, Ernest W., State Reports. Durham, N.C.: NTS Research Corp., July 1977.

Mason, Ward S. Problems of Measurement and the NIE Program. Washington, D.C.: National Institute of Education, 1973.

Murphy, Jerome T. State Education Agencies and Discretionary Funds; Grease the Squeaky Wheel. Lexington, Massachusetts: D.C. Heath and Company 1974.

Nixon, President Richard M., Economic Opportunity Message To Congress, February 19, 1969, as quoted in Williams, Walter and Evans, John W., "The Politics of Evaluation: The Case for Head Start," The Annals of the American Academy of Political and Social Science, September 1969, 385: 118-132.

Parlett, Malcom and Hamilton, David. "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs," (1975) in Glass, Gene V. (Ed.), Evaluation Studies Review Annual, 1976, 1: 140-157.

Ross, Lee and Cronbach, Lee J. (Ed.) "Handbook of Evaluation Research," Educational Researcher (essay review by a Task Force of the Stanford Evaluation Consortium). November 1976, 5: 10, 9-19.

Rossi, P. H. "Testing Success and Failure in Social Action" in Rossi, P. H. and Williams, W., Evaluating Social Programs: Theory, Practice, and Politics, New York: Seminar Press, 1972, 11-49.

Rossi, Peter H. and Williams, Walter (Ed.), Evaluating Social Programs. New York: Seminar Press, 1972.

Rossi, Peter H. and Wright, Sonia R. "Evaluation Research: An Assessment of Theory, Practice, and Politics," Evaluation Quarterly, February 1977.

Schenkel, Walter and Sieber, Sam D. The Design of Survey Questionnaires: A Case History Approach. Mimeographed. New York: Bureau of Applied Social Research, Columbia University, 1969.

Scriven, Michael. "Evaluation Perspectives and Procedures," in Popham, W. James (Ed.), Evaluation in Education, American Educational Research Association, 1974, 3-93.

Sieber, Sam D. "The Integration of Fieldwork and Survey Methods," American Journal of Sociology, May 1973, 28:6, 1335-1359.

Sieber, Sam D., and Lazarsfeld, Paul F. The Organization of Educational Research. Cooperative Research Project No. 1974 (USOE). New York: Bureau of Applied Social Research, Columbia University, 1966.

Sieber, Sam D.; Louis, Karen S.; and Metzger, L. The Use of Educational Knowledge, Vol. 1-2. New York: Bureau of Applied Social Research, Columbia University, 1972.

Stake, Robert E. "The Case Study Method in Social Inquiry," Educational Researcher, February 1978, 7: 2, 5-7.

Stearns, Marian S. and David, Jane L. A Proposal for Research: Longitudinal Study of Progress in Implementation Part I--Technical Proposal in response to RFP 77-95, as amended. Proposal No. URU-77-127 (revised). California: SRI International, November 1977.

Stenner, A. J. and Madey, D. L. Policy Implications Analysis, Exercise No. 1, "Design for the Evaluation of the State Capacity Building Program in Dissemination," Durham, N.C.: NTS Research Corp., 1977.

Stenner, A. Jackson. An Overview of Information Based Evaluation: A Design Procedure, Information Based Evaluation Series, Book 1. North Carolina: IBEX, Incorporated, 1974.

Strang, Ernest W. Building Capacity in Dissemination: Literature Review. North Carolina: NTS Research Corp., 1977.

Stufflebeam, Daniel L. "Alternative Approaches to Educational Evaluation: A Self-Study Guide for Educators," in Popham, W. James (Ed.), Evaluation in Education, American Educational Research Association, 1974, 95-144.

Trow, Martin. "Comment on Participant Observation and Interviewing: A Comparison." Human Organization 16 (Fall): 33-35, 1975.

Webb, Eugene J., Campbell, Donald T., Schwartz, Richard D., Sechrest, Lee. Unobtrusive Measures: Nonreactive Research in the Social Sciences Chicago: Rand McNally, 1966.

Weiss, Robert S. and Rein, Martin. "Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," Annals of the American Academy of Political and Social Science, 385 (September 1969).

Williams, Walter and Evans, John W., "The Politics of Evaluation: The Case for Head Start," Annals of the American Academy of Political and Social Science, 385 (September 1969), 118-132.

Willems, Edwin P. and Raush, Harold L., Naturalistic Viewpoints in Psychological Research. New York, N.Y.: Holt, Rinehart, and Winston, 1969.

Wolf, Robert L. and Tymitz, Barbara. "Toward More Natural Inquiry in Education," n.d. (mimeo)

Zelditch, M., Jr. "Some Methodological Problems of Field Studies," American Journal of Sociology, March 1962, 67: 566-76.

APPENDIX A. EVALUATION QUESTIONS FOR THE NTS STUDY

A. IS DISSEMINATION CAPACITY BEING BUILT?

1. What knowledge can be generated regarding the origins, objectives and plans of the capacity building projects?

   a. What are the similarities and differences within and across states in

      - Project plans?
      - Comprehensive State Dissemination Plans?

   b. How much progress or growth has there been in building capacity

      - Relative to program goals?
      - Relative to project objectives?
      - Relative to a baseline or other previous measure?

2. What knowledge can be generated concerning service capacity?

   a. What are the chief contextual elements within service capacity which characterize and differentiate projects? SEAs? States?

   b. What are the chief system characteristics within service capacity which characterize and differentiate projects? SEAs? States?

   c. On the basis of identified contextual elements within service capacity, can projects/SEAs/States be clustered into a limited number of types?

   d. On the basis of identified system characteristics within service capacity, can projects/SEAs/States be clustered into a limited number of types?

   e. On the basis of identified contextual clusters and identified system characteristic clusters, can a limited number of models be derived?

   f. What are the chief output variables within service capacity which characterize and differentiate projects? SEAs? States? Models?

3. What knowledge can be generated concerning user capacity?

   a. What knowledge can be generated concerning potential client capacity?

      (1) Who are the potential clients?

      (2) What are the perceptions of potential clients regarding service needs?

      (3) To what extent are potential clients aware of the service capacity of the project? SEA? States?

      (4) To what extent are potential clients knowledgeable of how to access/utilize the Service Capacity of the project? SEA? State?

   b. What knowledge can be generated concerning actual clientele?

      (1) Who are the actual users of the service capacity of the project? SEA? State?

      (2) What are the actual service capacity needs as expressed in requests of the project? SEA? State?

      (3) To what extent are the clients satisfied with the dissemination services provided by the project? SEA? State?

      (4) In what ways are the services utilized which are provided by the project? SEA? State?

4. How good a match is there between service capacity and user capacity?

   a. What are the relationships between service capacity and user capacity?

   b. To what extent are changes in service capacity and user capacity accommodated by changes in the other?

B. IS THE PROGRAM HAVING AN EFFECT?

1. What effects are the project having?

   a. What impact is each project having on the

      - SEA dissemination efforts?

      - State dissemination efforts?

b. What are the similarities and differences across projects in terms of these impacts?

2. Is dissemination capacity being institutionalized so that it can continue to function without external stimulus?

3. What effects are related to the federal role?

   a. What is the impact of different elements of the NIE program design and management on capacity building and institutionalization?

   b. Are there changes that should be made to facilitate the building and institutionalization of state capacity?

NTS Research Corporation is a private research and evaluation corporation working to improve education throughout the country. NTS performs contract work for federal, state, and local government agencies. The company conducts independent programs and research, funded through government and foundation contracts and corporate contributions. In addition, NTS provides test scoring, large scale data reduction, and analysis for local schools and private corporations. The company develops and markets test for worldwide distribution. Since its founding in 1973, NTS has successfully conducted over 100 major research and evaluation contracts. Annual sales volume exceeds 2 million. NTS is directed by a three member executive board. The corporate offices are located in Durham, North Carolina, near the Research Triangle Park. These offices support field staff members through out the country.